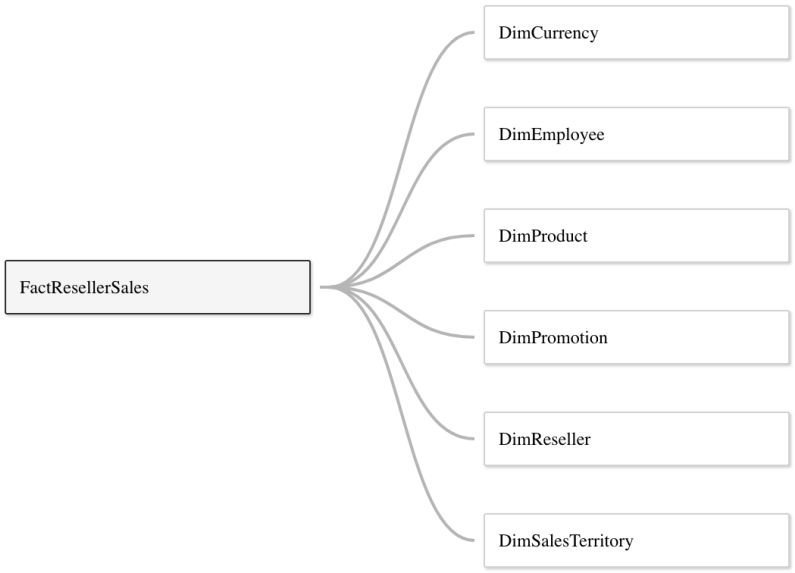
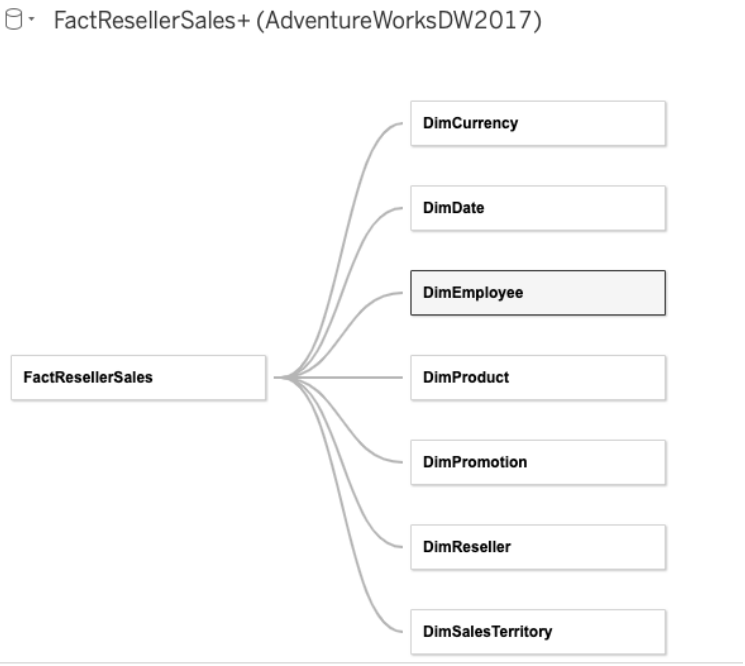
Tableau Analysis

This assignment adopt the sample Adventure Works Data Warehouse, connecting the tables FactResellerSales, DimCurrency, DimEmployee, DimProduct, DimPromotion, DimSeller, DimSalesTerritory.



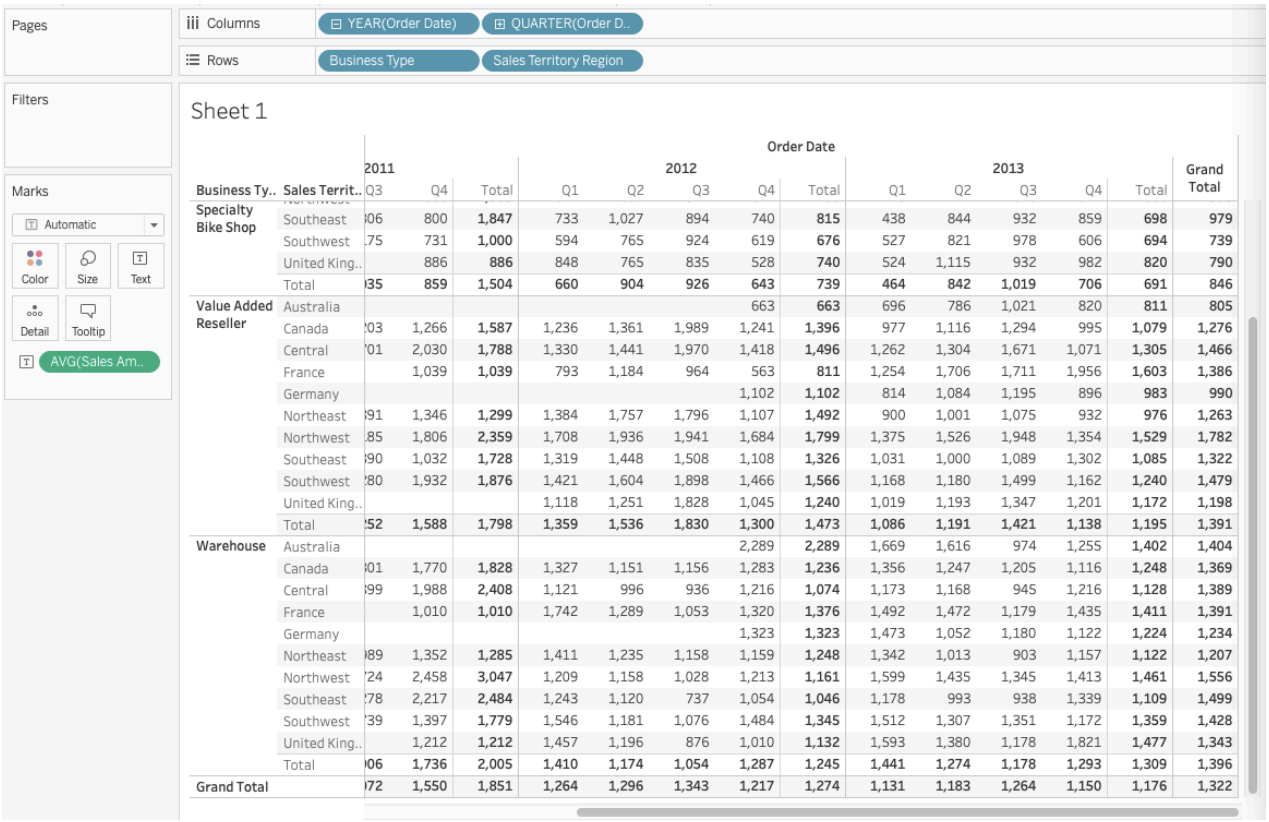
<u>Connected tables snapshot</u>: -



1. Cross Table Calculation and Aggregation:

(1). Create and display a cross table with ***"Business Type"*** firstly (from DimReseller) and ***"Sales Territory Region"*** secondly (from DimSalesTerritory) as rows, with ***Year of "Order Date"*** and ***Quarter of "Order Date"*** (discrete types) from FactResellerSales as columns; The value of the cells in this cross table is the ***average of "Sales Amount"*** from FactResellerSales;

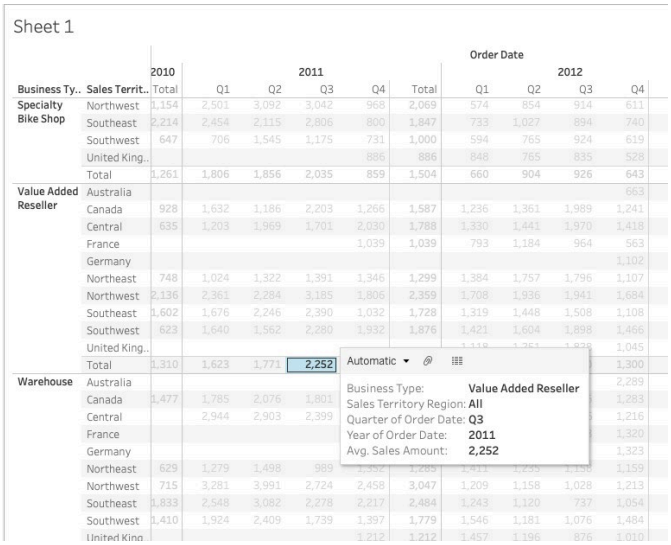
Besides the values in the cells, generate and display ***row grand totals***, ***column grand totals*** and ***all subtotals*** in the same cross table;



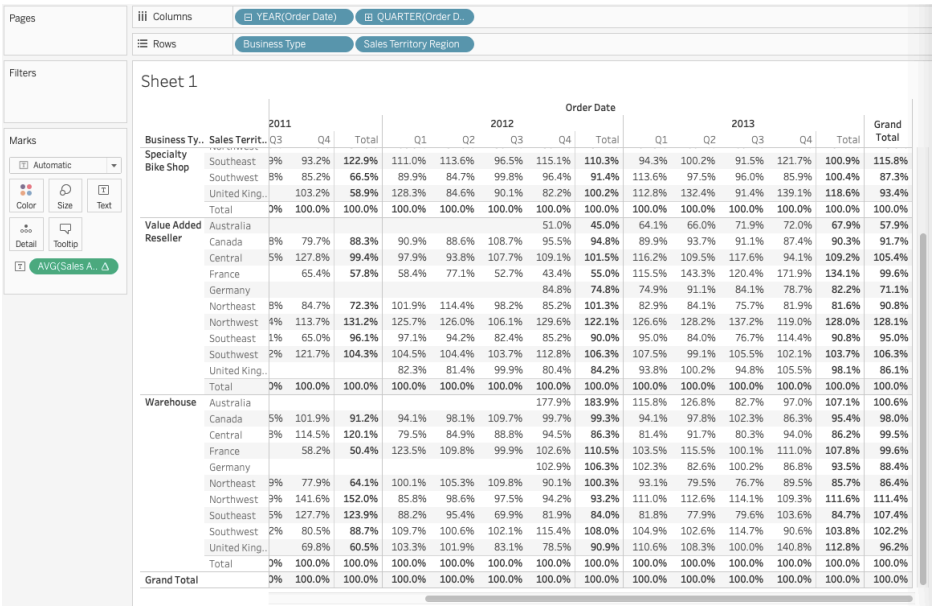
		Order Date													
		2011			2012					2013					Grand Total
Business Ty..	Sales Territ..	Q3	Q4	Total	Q1	Q2	Q3	Q4	Total	Q1	Q2	Q3	Q4	Total	
Specialty Bike Shop	Southeast	806	800	1,847	733	1,027	894	740	815	438	844	932	859	698	979
	Southwest	775	731	1,000	594	765	924	619	676	527	821	978	606	694	739
	United King..		886	886	848	765	835	528	740	524	1,115	932	982	820	790
	Total	835	859	1,504	660	904	926	643	739	464	842	1,019	706	691	846
Value Added Reseller	Australia							663	663	696	786	1,021	820	811	805
	Canada	203	1,266	1,587	1,236	1,361	1,989	1,241	1,396	977	1,116	1,294	995	1,079	1,276
	Central	701	2,030	1,788	1,330	1,441	1,970	1,418	1,496	1,262	1,304	1,671	1,071	1,305	1,466
	France		1,039	1,039	793	1,184	964	563	811	1,254	1,706	1,711	1,956	1,603	1,386
	Germany							1,102	1,102	814	1,084	1,195	896	983	990
	Northeast	391	1,346	1,299	1,384	1,757	1,796	1,107	1,492	900	1,001	1,075	932	976	1,263
	Northwest	385	1,806	2,359	1,708	1,936	1,941	1,684	1,799	1,375	1,526	1,948	1,354	1,529	1,782
	Southeast	390	1,032	1,728	1,319	1,448	1,508	1,108	1,326	1,031	1,000	1,089	1,302	1,085	1,322
	Southwest	380	1,932	1,876	1,421	1,604	1,898	1,466	1,566	1,168	1,180	1,499	1,162	1,240	1,479
	United King..				1,118	1,251	1,828	1,045	1,240	1,019	1,193	1,347	1,201	1,172	1,198
	Total	552	1,588	1,798	1,359	1,536	1,830	1,300	1,473	1,086	1,191	1,421	1,138	1,195	1,391
Warehouse	Australia							2,289	2,289	1,669	1,616	974	1,255	1,402	1,404
	Canada	801	1,770	1,828	1,327	1,151	1,156	1,283	1,236	1,356	1,247	1,205	1,116	1,248	1,369
	Central	999	1,988	2,408	1,121	996	936	1,216	1,074	1,173	1,168	945	1,216	1,128	1,389
	France		1,010	1,010	1,742	1,289	1,053	1,320	1,376	1,492	1,472	1,179	1,435	1,411	1,391
	Germany							1,323	1,323	1,473	1,052	1,180	1,122	1,224	1,234
	Northeast	889	1,352	1,285	1,411	1,235	1,158	1,159	1,248	1,342	1,013	903	1,157	1,122	1,207
	Northwest	724	2,458	3,047	1,209	1,158	1,028	1,213	1,161	1,599	1,435	1,345	1,413	1,461	1,556
	Southeast	778	2,217	2,484	1,243	1,120	737	1,054	1,046	1,178	993	938	1,339	1,109	1,499
	Southwest	739	1,397	1,779	1,546	1,181	1,076	1,484	1,345	1,512	1,307	1,351	1,172	1,359	1,428
	United King..		1,212	1,212	1,457	1,196	876	1,010	1,132	1,593	1,380	1,178	1,821	1,477	1,343
	Total	806	1,736	2,005	1,410	1,174	1,054	1,287	1,245	1,441	1,274	1,178	1,293	1,309	1,396
Grand Total		772	1,550	1,851	1,264	1,296	1,343	1,217	1,274	1,131	1,183	1,264	1,150	1,176	1,322

(2). What is the subtotal of sales amount from ***Value Added Reseller*** in ***2011 Q3***?

Answer: *The subtotal of average of sales amount from Value Added Reseller in 2011 Q3 is* ***2,552****.*



(3). Based on the prior cross table, generate and display ***quick table calculation*** with ***percentage of total***, using ***pane (down)***;

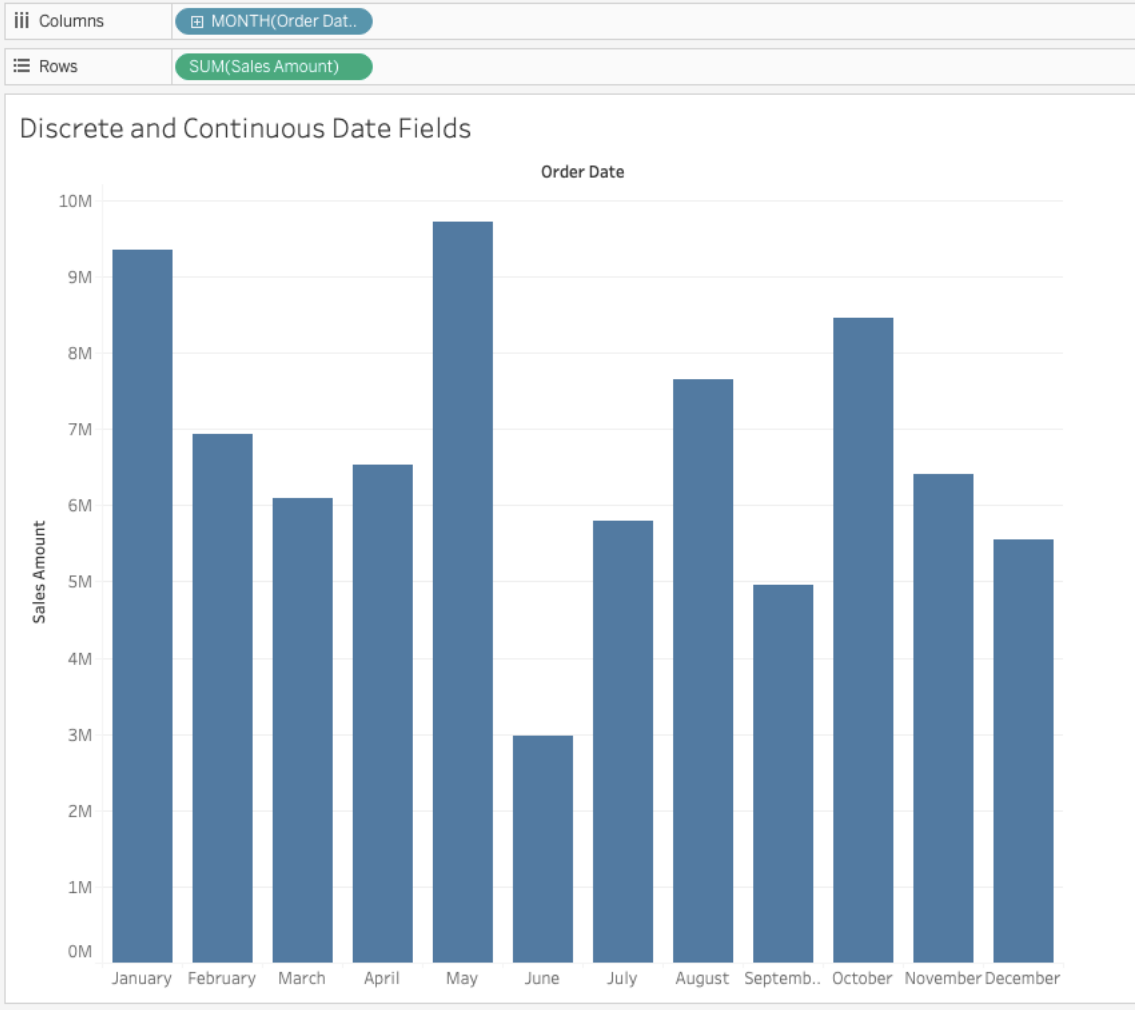


(4). Based on the cross table from (3), Why do we have ***100%*** as the subtotal value at each pane level?

Answer*: Each pane is representing* "Business Type" attribute for each quarter. As we are computing based on each pane in the downward direction, hence the subtotal is calculated as 100% at each pane level.

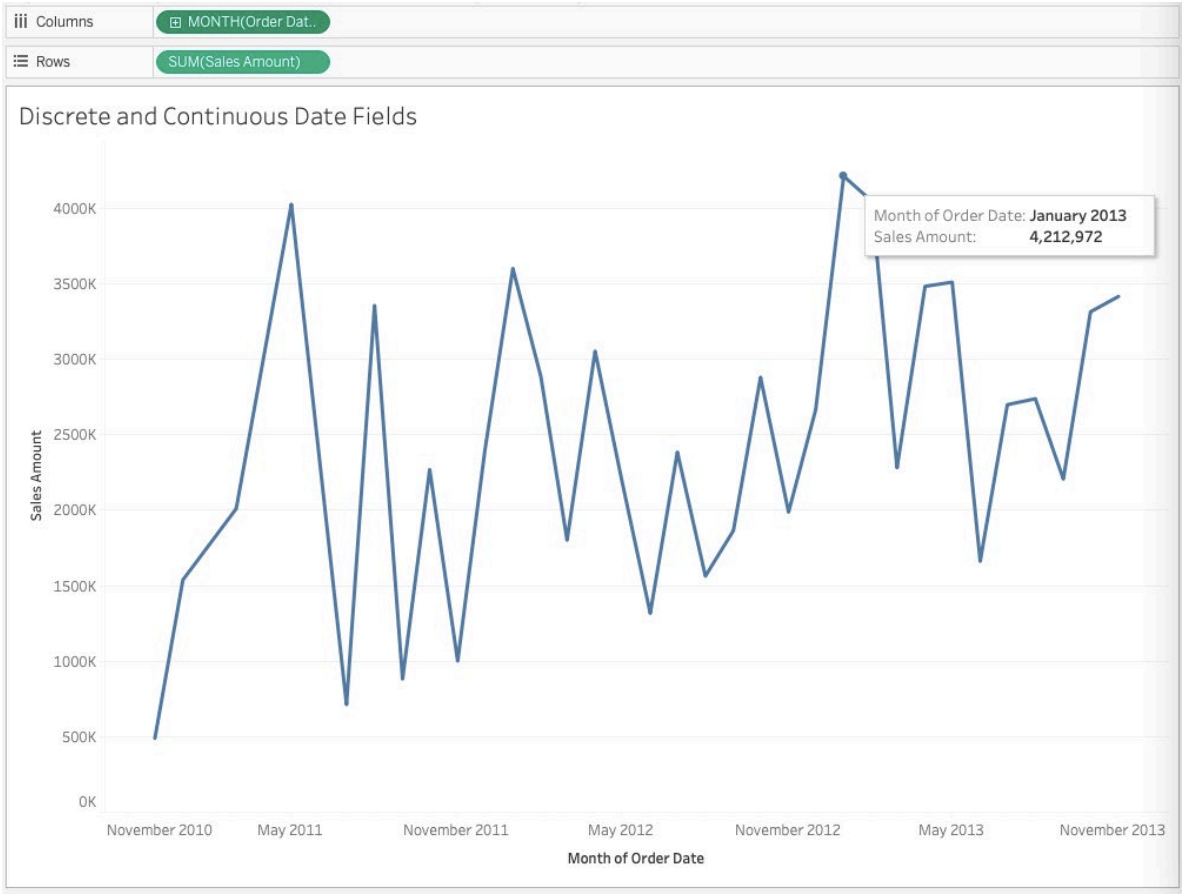
2. Discrete and Continuous Date Fields:
(1). Create and display a bar chart, with ***sum of "Sales amount"*** as rows, and **discrete** type ***month of "Order Date"*** from FactResellerSales table. Which month overall has the lowest ***sum of sales amount***?

Answer: ***June*** *month has overall the lowest sum of sales amount*



(2). Create and display a line chart, with ***sum of "Sales Amount"*** as rows, and **continuous** type ***month of "Order Date"*** from FactResellerSales table. Which month and year has the highest sum of sales amount?

Answer: ***January of 2013*** *has the highest sum of sales amount.*



(3). When do we usually use continuous date field, and when do we use discrete date field?

Answer: *Discrete means individually separate and Continuous means unbroken, without interruption. Discrete date field is used when my analysis requires to have distinct marks that can be sorted, whereas Continuous date field is used to look at a trend over a continuous time period. Continuous date fields cannot be sorted.*
Discrete data will generate rows and columns, whereas continuous data generates axes.

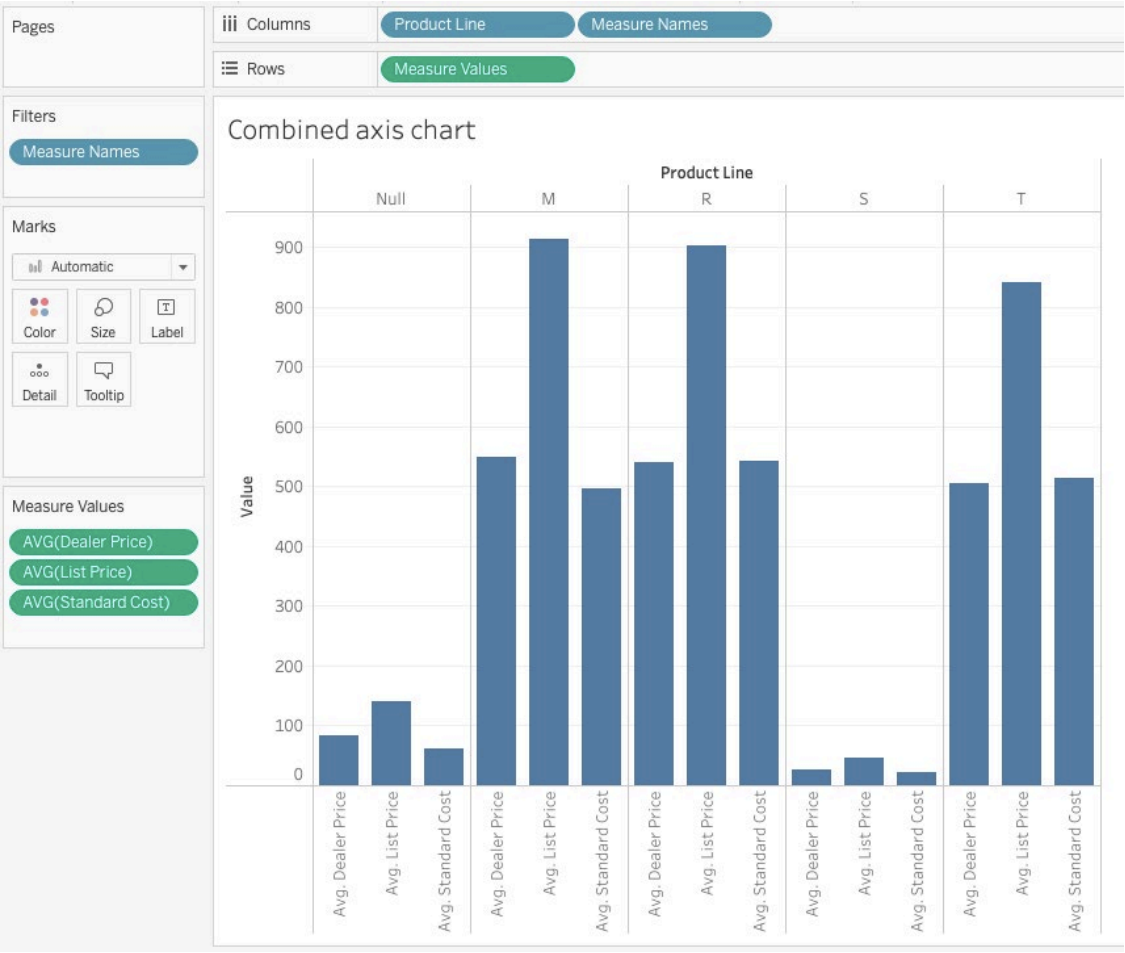
3. Dual axis and Combined axis chart:
(1). Compare the ***average of "Sick Leave Hours"*** (from DimEmployee) and the ***average of "Vacation Hours"*** (from DimEmployee) using a dual axis chart, with ***"Department Name" (from DimEmployee)*** as columns; synchronize the two axis and display the chart;

(2). Based on the dual axis chart, which departments have higher average sick leave hours than average vacation hours?

Answer: *Departments which have higher average sick leave hours than average vacation hours are: -*

- *Engineering*
- *Executive*
- *Research and Development*
- *Sales*
- *Tool Design*

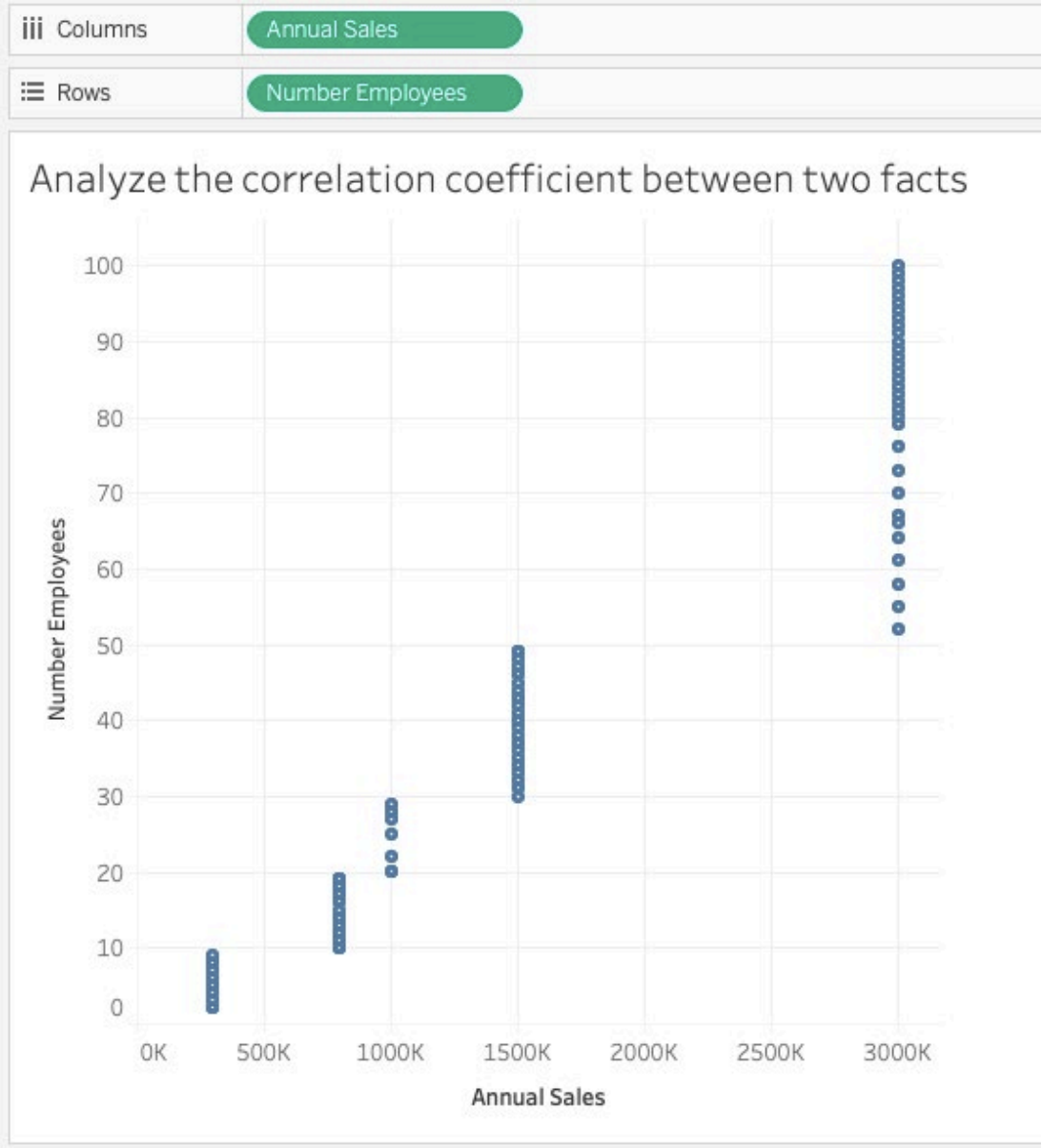
(3). Generate and display a combined axis chart, with ***"Product Line"*** (from DimProduct) as columns, and ***average of "Dealer Price"***, ***average of "List Price"***, and ***average of "Standard Cost"*** (all from DimProduct) as rows;



(4). Based on the prior combined axis chart, which product line has the highest average "List Price", and which product line has the lowest average "List Price"?

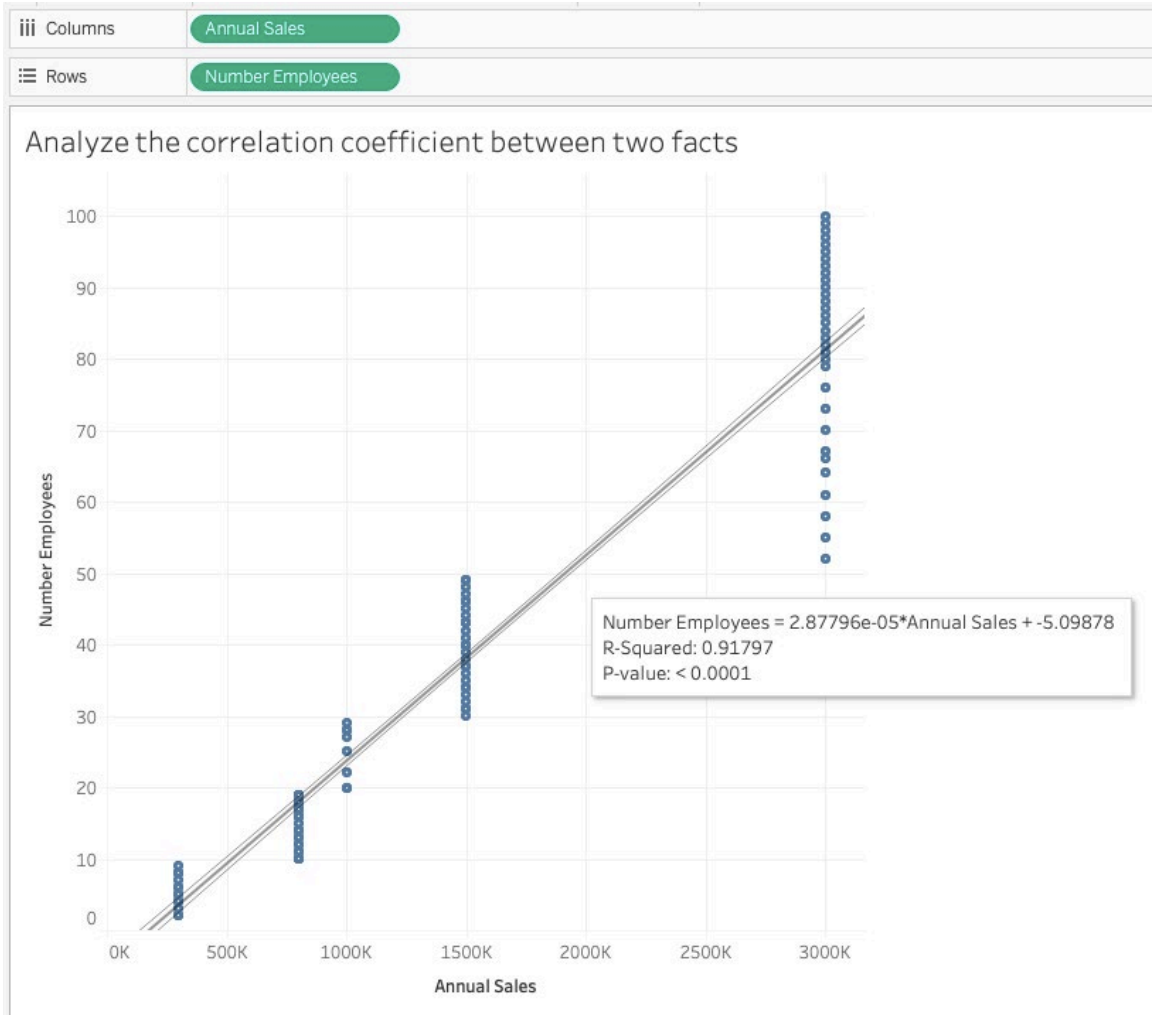
Answer: The *product line* ***M has the highest average "List Price"*** *and the product line* ***S has the lowest average "List Price"****.*

4. Analyze the correlation coefficient between two facts:
(1). Generate and display a scatter plot with facts ***"Annual Sales"*** and ***"Number Employees"*** from DimReseller; (hint: uncheck the aggregated measures)



Marks count - 701

(2). Generate and display ***a linear trend line*** on top of the scatter plot with facts “Annual Sales” and “Number Employees”, enable the confidence interval of the linear line.

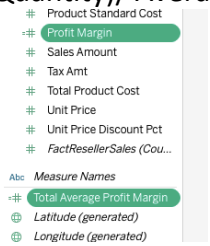


(3). Based on the linear trend line, what is the R square value? What is the p value? What is the direction of the correlation coefficient between Annual Sales” and “Number Employees”? Are the correlation coefficient significant?

Answer: *Base on the linear trend line: -*

- *R-squared value:* ***0.91797***
- *P-value: <* ***0.0001***
- *The direction of the correlation coefficient between Annual Sales and Number Employees:* ***upward or positively linear***
- *As the p-value is less than 0.5, the correlation coefficient between the two* ***is significant****.*

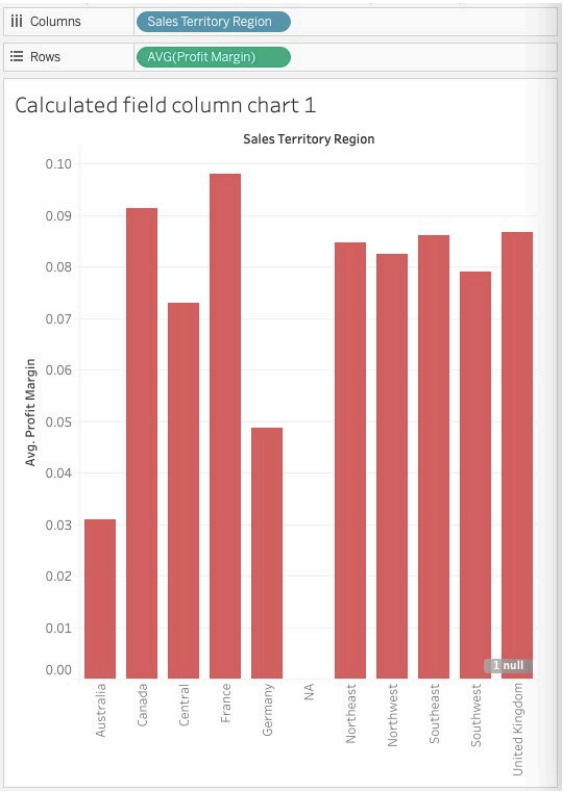
5. Create Calculated Fields:
(1). Create and display calculated fields:
a. ***Profit Margin*** = ((Unit Price – Product Standard Cost)*Order Quantity)/Sales Amount;
b. ***Total Average Profit Margin*** = Average ((Unit Price – Product Standard Cost)*Order Quantity)/ Average (Sales Amount);

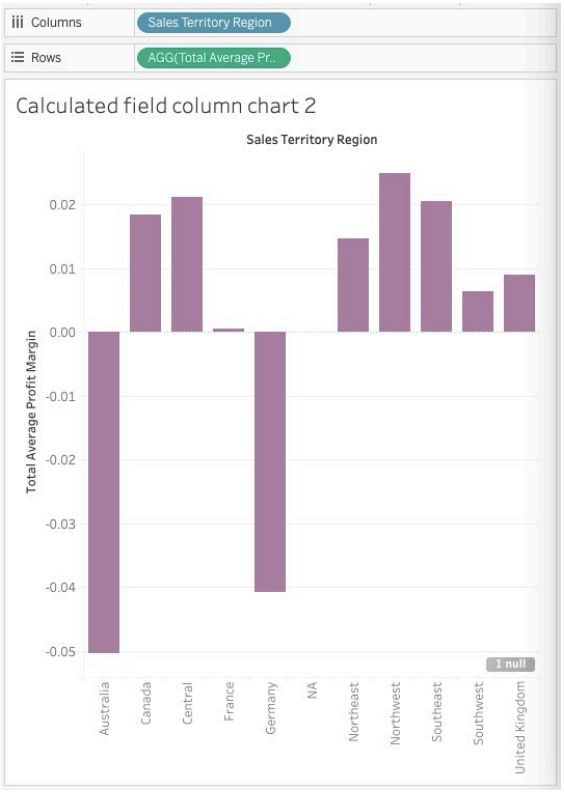


(2). Generate and display two column charts, with ***average of "Profit Margin"*** and "***Sales Territory Region"*** (from DimSalesTerritory) in one chart, and "Total Average Profit Margin" and "Sales Territory Region" in another chart; Why are these two charts look different?

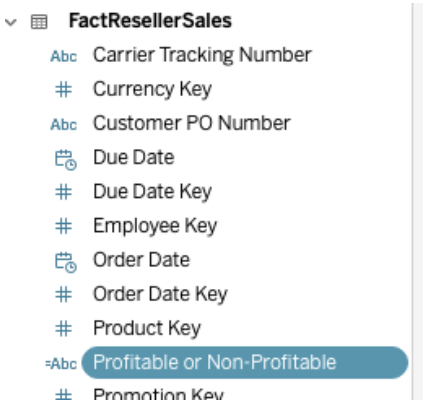
Chart 1: -
Average (Profit Margin = (([Unit Price] - [Product Standard Cost]) * [Order Quantity]) / [Sales Amount]**) for a region**; First one is the average of all the profit margins, where profit margin is profit for each divided by sales amount of each.

Chart 2: - (Total Profit Margin = AVG(([Unit Price] - [Product Standard Cost]) * [Order Quantity]) / AVG([Sales Amount])**) for a region**; Second one is average of Profits divided by average of sales amount.

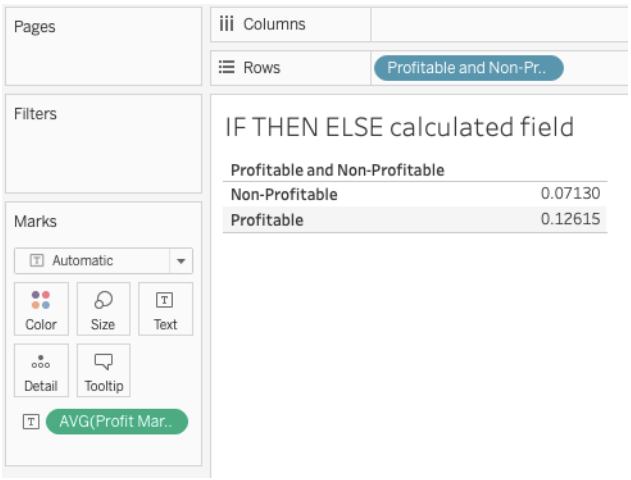




(3). Create another calculated field using IF THEN ELSE statement:
Create and display and new calculated field ***"Profitable"***: If "Unit Price" – "Product Standard Cost" is greater than 25, then we assign a value "Profitable"; otherwise, we assign the value "Nonprofitable";



(4). Create and display a cross table, with the levels from ***"Profitable"*** as rows, and ***average of "Profit margin"*** as cell values in this cross table;



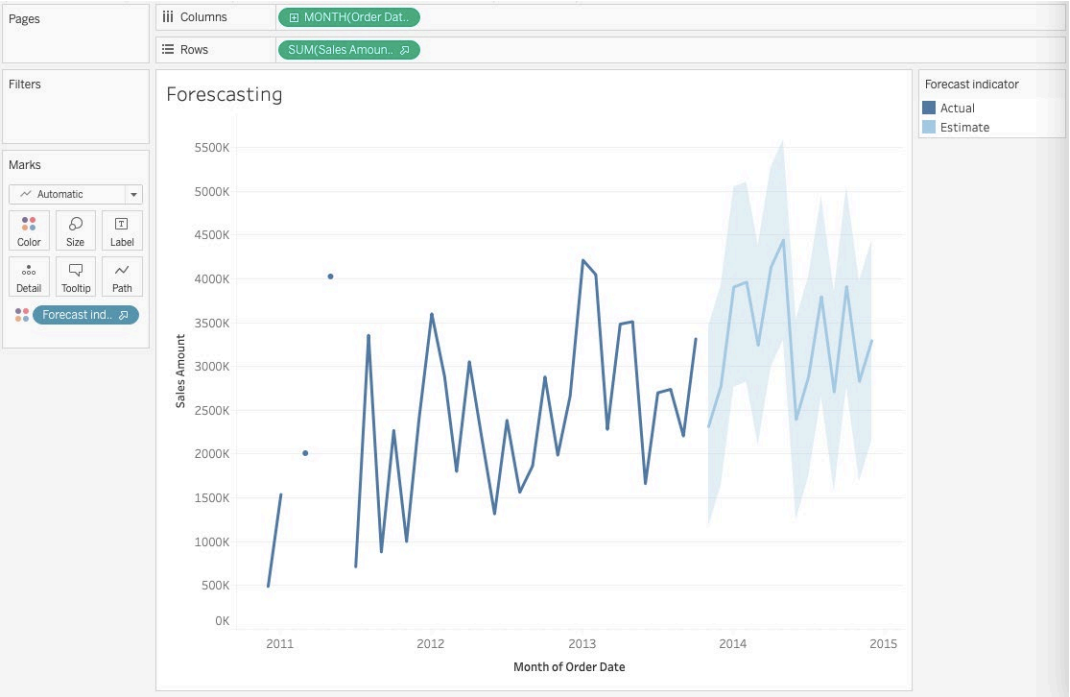
(5). What are the ***average profit margins*** for nonprofitable and profitable categories respectively?

Answer: *The average profit margins for respective categories are as follows: -*

- *Profitable:* ***0.12615***
- *Non-Profitable:* ***0.07130***

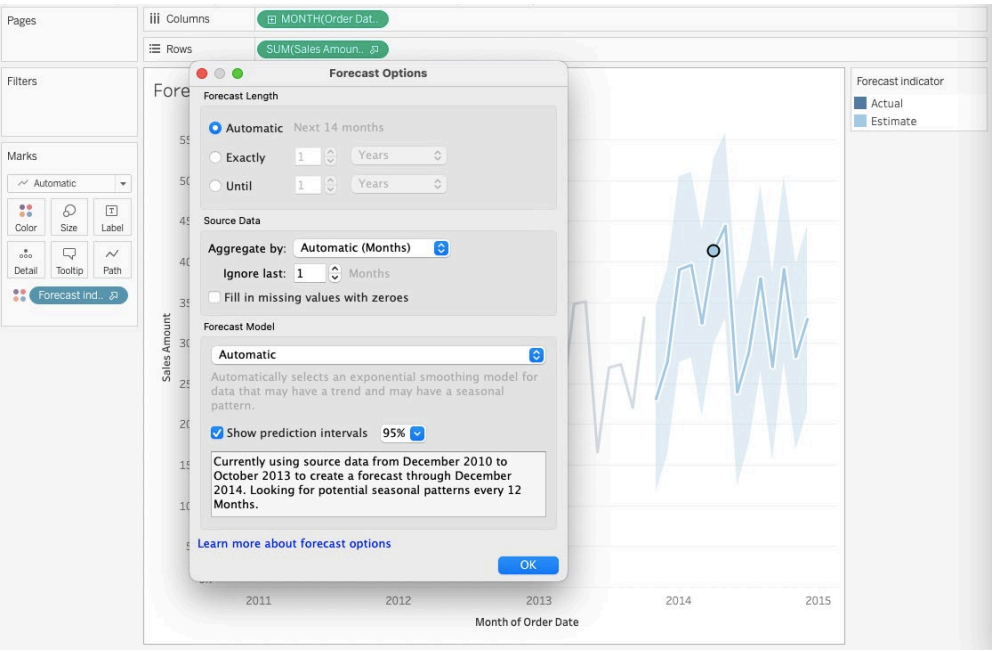
6. Forecasting and Clustering:

(1). Generate and display a forecasting line predicting sum of “Sales Amount” on the basis of month of “Order Date” (continuous) from FactResellerSales table;



(2). Use default settings from forecast options, how many months’ sum of sales are predicted in the model?

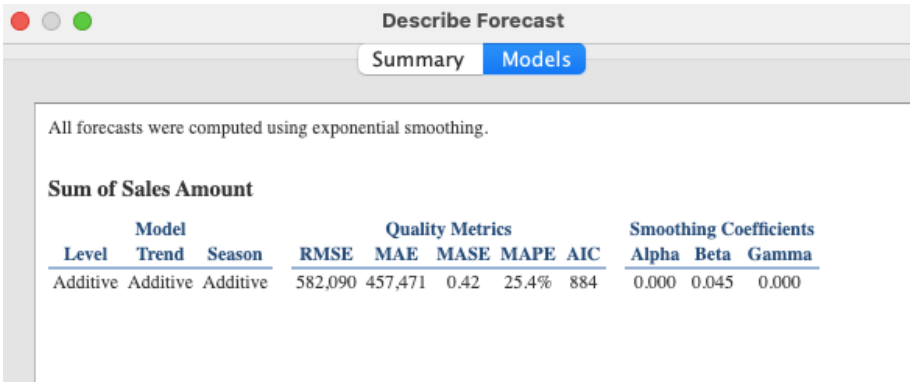
Answer: *Sum of sales have been predicted for next **14 months** in the forecasting model.*



(3). From the summary of the models, what are the quality metrics of the forecasting model?

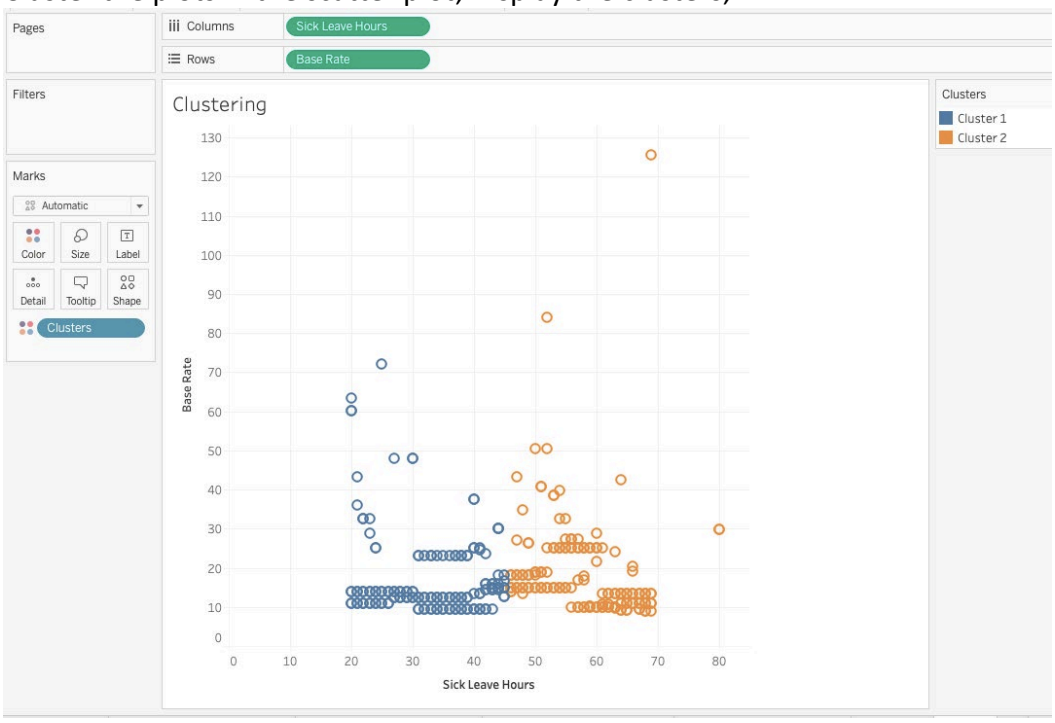
Answer: *The quality metrics of the forecasting model are: -*

Quality Metrics				
RMSE	MAE	MASE	MAPE	AIC
582,090	457,471	0.42	25.4%	884



Model			Quality Metrics					Smoothing Coefficients		
Level	Trend	Season	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Additive	Additive	Additive	582,090	457,471	0.42	25.4%	884	0.000	0.045	0.000

(4). Generate a scatter plot with “Sick Leave Hours” and “Base Rate” from Dimemployee; Cluster the plots in the scatter plot; Display the clusters;



(5). Use the default setting, how many clusters are generated?

Answer: *By default, two clusters have been generated.*

(6). Change the cluster number to 3, display the clustered scatter plot, how many items are in each cluster?

Answer: *Number of items in each cluster are as follows: -*

- *Cluster 1: **95** items*
- *Cluster 2: **106** items*
- *Cluster 3: **95** items*